



(19) RU (11) 2 138 076 (13) C1
(51) МПК⁶ G 06 F 17/30

РОССИЙСКОЕ АГЕНТСТВО
ПО ПАТЕНТАМ И ТОВАРНЫМ ЗНАКАМ

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ПАТЕНТУ РОССИЙСКОЙ ФЕДЕРАЦИИ

(21), (22) Заявка: 98116879/09, 14.09.1998

(24) Дата начала действия патента: 14.09.1998

(46) Дата публикации: 20.09.1999

(56) Ссылки: WO 98/20434, 14.05.98, US 5748954 A, 05.05.98. US 5576954 A, 19.11.96.
Управляющие вычислительные машины в АСУ технологическими процессами/Под ред. Т.Харрисона. - М.: Мир, 1975, с.48 - 52, фиг.1.11.

(98) Адрес для переписки:
115446, Москва, Коломенский пр-д 1а, ЗАО
"МедиаЛингва", Ашманову И.С.

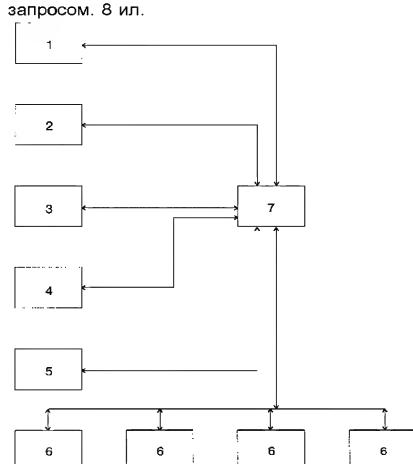
(71) Заявитель:
Закрытое акционерное общество
"МедиаЛингва"

(72) Изобретатель: Серебренников О.А.,
Ашманов И.С., Иванов А.П., Пашко Д.А.

(73) Патентообладатель:
Закрытое акционерное общество
"МедиаЛингва"

(54) СИСТЕМА ПОИСКА ИНФОРМАЦИИ В КОМПЬЮТЕРНОЙ СЕТИ

(57) Реферат:
Изобретение относится к вычислительной технике. Технический результат заключается в расширении функциональных возможностей и повышении эффективности использования оборудования, который достигается за счет того, что система поиска информации в компьютерной сети содержит устройство сбора и обработки документов и память документов и адресов. Система содержит также устройства управления абонентов, устройства индексирования документов и индексную память. Устройство поиска выполнено с возможностью приема поисковых индексных образов и запросов на поиск от устройств управления абонентов, преобразования запросов на поиск в индексный образ, сравнения с индексными образами индексной памяти, выбора адресов документов, соответствующих индексному образу абонентского запроса, ранжирования выбранных адресов по степени соответствия индексному образу абонентского запроса и формирования списка найденных документов с указанием их адресов в компьютерной сети в соответствии с указанным абонентским



ФИГ. 1

RU 2 138 076 C1

RU ? 1 3 8 0 7 6 C 1



(19) RU (11) 2 138 076 (13) C1
(51) Int. Cl.⁶ G 06 F 17/30

RUSSIAN AGENCY
FOR PATENTS AND TRADEMARKS

(12) ABSTRACT OF INVENTION

(21), (22) Application: 98116879/09, 14.09.1998

(24) Effective date for property rights: 14.09.1998

(46) Date of publication: 20.09.1999

(98) Mail address:
115446, Moskva, Kolomenskij pr-d 1a, ZAO
"MediaLingva", Ashmanov I.S.

(71) Applicant:
Zakrytoe aktsionernoe obshchestvo "MediaLingva"

(72) Inventor: Serebrennikov O.A.,
Ashmanov I.S., Ivanov A.P., Pashko D.A.

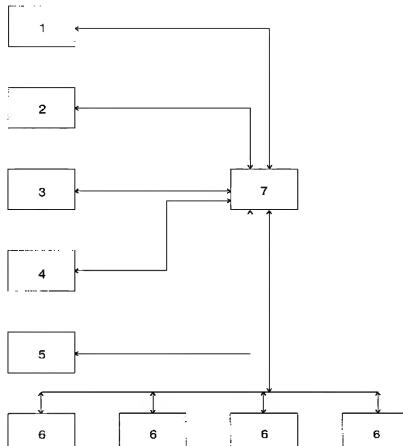
(73) Proprietor:
Zakrytoe aktsionernoe obshchestvo "MediaLingva"

(54) DATA RETRIEVAL SYSTEM IN COMPUTER NETWORK

(57) Abstract:

FIELD: computer engineering. SUBSTANCE: system has document acquisition and processing device as well as document and address memory. It also has subscriber control devices, document indexing devices, and index memory. Data retrieval device is made for receiving search index images and search requests coming from subscriber control devices, converting search requests to index image, comparing it with index images of index memory, retrieving addresses of documents corresponding to index image of subscriber request, ranking chosen addresses according to their correspondence to index image of subscriber request, and forming list of found documents including their addresses in computer network according to mentioned subscriber request. EFFECT: enlarged functional capabilities and improved efficiency of hardware use. 8 dwg

RU ? 1 3 8 0 7 6 C1



ФИГ. 1

RU
2 1 3 8 0 7 6 C1

RU ? 1 3 8 0 7 6 C 1

Изобретение относится к области вычислительной техники и может быть использовано для поиска информации в компьютерной сети, например, Интернет, по абонентскому запросу на естественном и/или формальном языке.

Известны устройство и способ поиска релевантных документов (см. патент США N 5576954, опубл. 19.11.96 г.), осуществляющее сравнение различных документов и ранжирование их по степени соответствия.

Известная система предусматривает обращение в сеть только на формальном языке.

Известен способ и устройство поиска документов в памяти, их классификации и размещения в определенном порядке (см. патент США N 5748954, опубл. 05.05.98 г.).

Известная система предусматривает обращение в компьютерную сеть только на формальном языке.

Известна система для связи абонентов в Интернете, сбора и отображения информации (см. заявка PCT WO 98/20434, опубл. 14.05.98 г.). Известная система выбрана как наиболее близкий аналог к патентуемой.

Известная система позволяет обращаться в компьютерную сеть только на формальном языке.

Задачей патентуемого изобретения является обеспечение возможности обращения в компьютерную сеть с запросом на поиск документов любым образом, в т. ч. на естественном языке.

Достигаемый результат состоит в расширении функциональных возможностей и повышении эффективности использования оборудования системы поиска информации в компьютерной сети.

Результат достигается тем, что в систему поиска информации в компьютерной сети, содержащую устройство сбора и обработки документов с возможностью приема документов из серверов сети, выделения в полученных документах адресов, приема документов в соответствии с выделенными адресами, обмена документами и ссылками в память документов и адресов, анализа наличия документов и адресов в памяти документов и адресов, изменения и удаления документов из памяти документов и адресов по результатам анализа адресов, память документов и адресов с возможностью приема и выдачи документов и адресов устройства управления абонентов с возможностью формирования запросов на поиск на формальном и/или естественном языке, приема и удаления документов и адресов, заголовков, атрибутов, аннотаций документов, в отличие от ближайшего аналога введены устройство индексирования документов с возможностью формирования индексов документов, выделения заголовков и аннотаций документов и их частей, выделения атрибутов документов, формирования индексного формата документа, формирования индексного образа документов путем компоновки индексного формата, адресов документов и атрибутов документов, приема адресов изменяемых или удаляемых документов, удаления индексных образов документов, передачи поступивших адресов и сформированных индексных образов в индексную память, индексную память с возможностью приема адресов и

индексных образов, передачи индексных образов в устройство поиска, устройство поиска с возможностью приема поисковых индексных образов и запросов на поиск от устройств управления абонентов, преобразования запросов на поиск в индексный образ, сравнения с индексными образами индексной памяти, выбора адресов документов, соответствующих индексному образу абонентского запроса, ранжирования выбранных адресов по степени соответствия индексному образу абонентского запроса и формирования списка найденных документов с указанием их адресов в компьютерной сети в соответствии с указанным абонентским запросом.

Конкретный, но не ограничивающий патентуемое изобретение, пример выполнения системы приведен на фиг. 1-8, где указаны устройство 1 сбора и обработки документов, память 2 документов и их

адресов, устройство 3 индексирования документов, индексная память 4, устройство 5 поиска, устройство 6 управления абонентов, устройство 7 межсистемного обмена, устройство 1 содержит блок 8 микропроцессорного управления, буферную память 9 документов, блок 10 выборки адресов документов, таймер 11, блок 12 буферной памяти адресов удаляемых документов, буферную память 13 документов, коммутатор 14, блок 8 микропроцессорного управления содержит память 15, узел 16 управления и обработки, узел 17 ввода-вывода, память 2 содержит блок 18 микропроцессорного управления, блок 19 памяти документов, блок 20 памяти адресов, таймер 21, коммутатор 22, блок 3 индексирования содержит блок 23

микропроцессорного управления, блок 24 выделения заголовков документов, блок 25 выделения аннотации документов, блок 27 выделения атрибутов документов, блок 28 формирования индексного образа документа, таймер 29, буферную память 30 адресов, буферную память 31 адресов удаляемых или изменяемых документов, коммутатор 32, индексная память 4 содержит блок 33 микропроцессорного управления, блок 34 памяти индексов, блок 35 памяти атрибутов, блок 36 памяти адресов, коммутатор 37, устройство 5 поиска содержит блок 38 микропроцессорного управления, блок 39 фиксации запроса, блок 40 фиксации атрибутов, буферную память 41 индексов, блок 42 сравнения, память 43 адресов, блок 44 ранжирования, блок 45 формирования списка, устройство 6 управления абонента содержит блок 46 микропрограммного управления, пульт 47 управления и отображения, блок 48 формирования запроса, блок 49 фиксации ответа, блок 50 анализа.

Все узлы, входящие в состав системы, могут быть реализованы с помощью программно-аппаратных средств, основанных на микросхемах, например, фирмы Intel. Система работает следующим образом. Устройство 1 производит опрос серверов компьютерной сети по известным адресам документов, производит копирование документов и их адресов в память 2, контролирует изменения документов и осуществляет удаление документов и адресов из памяти 2, если это необходимо.

Работа устройства осуществляется под

R
U
2
1
3
8
0
7
6
C
1

? 1 3 8 0 7 6 C 1

управлением блока 8, в котором функции контроля, анализа и обработки осуществляют узел 16 по программе памяти 15, а обмен информацией осуществляется через узел 17.

Документ из выбранного сервера считывается через узел 17 в память 9, из которой блоком 10 выбираются адреса связанных документов и передаются в память 13. Адреса из памяти 13 и памяти 2 сравниваются в блоке 8 и, если обнаруживается совпадение, что указывает на наличие документа в памяти 2, то передача информации из памяти 9 через коммутатор 14 в память 2 не производится в следующем цикле работы памяти 9 эта информация просто стирается. Если адрес не обнаружен в памяти 2, то осуществляется передача документа из памяти 9 через коммутатор 14, узел 17 и устройство 7 в память 2. Устройство 1 опрашивает серверы компьютерной сети и, если обнаруживает адрес документа, подлежащего удалению или изменению, то выбирает его из памяти 2 и производит соответствующую процедуру удаления или замены.

Процедура замены или удаления документов может быть также инициирована таймером 11. По сигналу таймера производится выборка адреса документа из памяти 2, который через коммутатор 14 передается в узел 17 ввода-вывода. Далее производится проверка наличия документа по заданному адресу. Если оказывается, что документ был обновлен, производится закачка новой версии документа, а предыдущее удаляется из памяти 2. Если документ по заданному адресу отсутствует, информация о нем удаляется из памяти 2.

Аналогичную процедуру осуществляет устройство 1, если при повременном сканировании таймером 11 адресов, выбираемых из памяти 13, обнаруживает адрес, указывающий на изменяемый или удаляемый документ.

Память 2 работает под управлением блока 18, имеющего структуру, аналогичную блоку 8. Принимаемые документы и адреса хранятся в блоках соответственно 19, 20 и, при необходимости, передаются через коммутатор 22 и узел ввода-вывода блока 18 в устройство 7 и далее по требованию. Таймер 21 позволяет выделить адрес и соответствующий документ, которые необходимо удалить или изменить.

Любой поступающий в систему документ подвергается некоторым преобразованиям для формирования индексного образа. Это осуществляется в устройстве 3, работающем под управлением блока 23, структура которого идентична структуре блока 8. По команде с таймера 29 узел ввода-вывода блока 23 производит из памяти 2 выборку новых адресов и документов. При этом исходные тексты документов передаются на входы блоков 24-27, адреса документов - в блок 30, а адреса удаляемых документов - в блок 31. Из документов выделяются заголовки блоком 24, аннотации - блоком 25, формируются индексные форматы блоком 26 путем статистической обработки документов и атрибуты документов блоком 27. Эти компоненты в блоке 28 объединяются в поисковый индексный образ и совместно со ссылкой документа из памяти 30 передаются через коммутатор 32, узел ввода-вывода

блока 23 и устройство 7 в память 4. Если обнаруживаются ссылки удаляемых или изменяемых документов, то они поступают в память 31 и далее в память 4, где осуществляется выборка соответствующих индексов и их последующая обработка. Индексы могут быть подвергнуты обработке и по результатам временного сканирования с помощью таймера 29.

В памяти 4 хранятся индексные образы документов в блоке 34, атрибуты - в блоке 35, соответствующие адреса документов в блоке 36, информация из которых может передаваться через коммутатор 37. Работа памяти 4 осуществляется под контролем и управлением блока 33, имеющего структуру, аналогичную структуре блока 8.

Работа устройства 5 осуществляется под управлением блока 38, имеющего структуру, аналогичную структуре блока 8. Через узел ввода-вывода блока 38 абонентский запрос от соответствующего устройства 6 поступает в блок 39, а в блок 40 и память 41 поступают из памяти 4 соответственно атрибуты документов и индексные образы документов. В блоке 42 осуществляется сравнение запроса абонента и, в случае необходимости, дополнительных атрибутов, с индексными образами документов. Результаты сравнения фиксируются в памяти 43 и ранжируются в блоке 44 с использованием информации от памяти 41. Результаты ранжирования фиксируются блоком 45 и передаются в устройство 6 соответствующего абонента.

В устройстве 6 работа управляется и контролируется блоком 46, структура которого идентична структуре блока 8. Кроме того, осуществляется ручное управление и контроль с помощью пульта 47. Запрос абонента формируется в блоке 48, ответ из устройства 5 фиксируется в блоке 49 и с помощью блока 50 осуществляется анализ ответа и, если требуется, формирует дополнительные сведения в блоке 48.

Таким образом, система позволяет проводить сбор и обновление информации о документах, хранящихся в компьютерной сети, сохранять информацию о документах в специализированной базе данных индексного хранилища, производить поиск документов в компьютерной сети на основе естественного или формального запроса пользователя, а также по набору различных атрибутов документов, ранжировать список найденных документов по степени соответствия запроса пользователя, получать информацию о документах сети без непосредственного обращения к ним.

Это обеспечивает расширение функциональной возможности и повышение эффективности использования оборудования системы поиска информации в компьютерной сети.

Формула изобретения:

Система поиска информации в компьютерной сети, содержащая устройство сбора и обработки документов с возможностью приема документов из серверов сети, выделения в полученных документах адресов, приема документов в соответствии с выделенными адресами, обмена документами и ссылками с памятью документов и адресов, анализа наличия документов и адресов в памяти документов и адресов, изменения и удаления документов

R U ? 1 3 8 0 7 6 C 1

из памяти документов и адресов по результатам анализа адресов, память документов и адресов, изменения и удаления документов из памяти документов и адресов по результатам анализа адресов, память документов и адресов с возможностью приема и выдачи документов и адресов, устройства управления абонентов с возможностью формирования запросов на поиск на формальном и/или естественном языке, приема и удаления документов и адресов, заголовков, атрибутов, аннотаций документов, отличающаяся тем, что в нее введены устройство индексирования документов с возможностью формирования индексов документов, выделения заголовков и аннотаций документов и их частей, выделения атрибутов документов, формирования индексного формата документа, формирования индексного образа документов путем компоновки индексного формата, адресов документов и атрибутов

документов, приема адресов изменяемых или удаляемых документов, удаления индексных образов документов, передачи поступивших адресов и сформированных индексных образов в индексную память, индексная память с возможностью приема адресов и индексных образов, передачи индексных образов в устройство поиска, устройство поиска с возможностью приема поисковых индексных образов и запросов на поиск от устройств управления абонентов, преобразования запросов на поиск в индексный образ, сравнения с индексными образами индексной памяти, выбора адресов документов, соответствующих индексному образу абонентского запроса, ранжирования выбранных адресов по степени соответствия индексному образу абонентского запроса и формирования списка найденных документов с указанием их адресов в компьютерной сети в соответствии с указанным абонентским запросом.

20

25

30

35

40

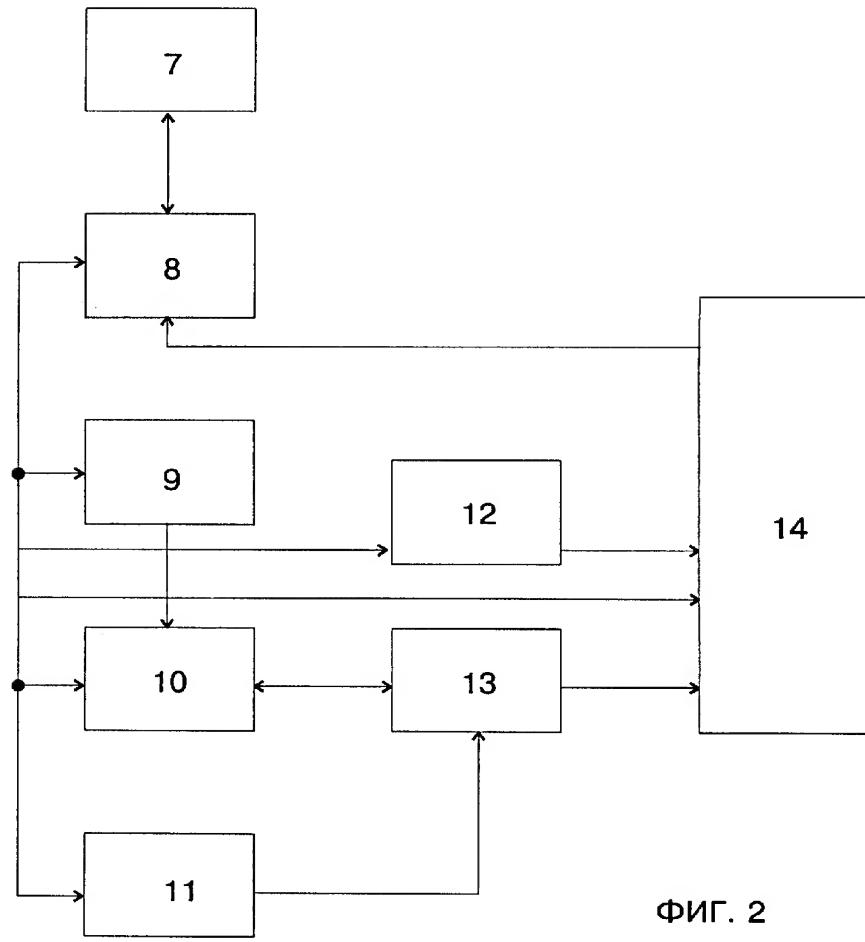
45

50

55

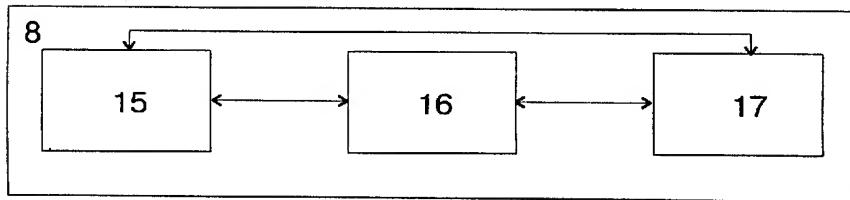
60

R U 2 1 3 8 0 7 6 C 1



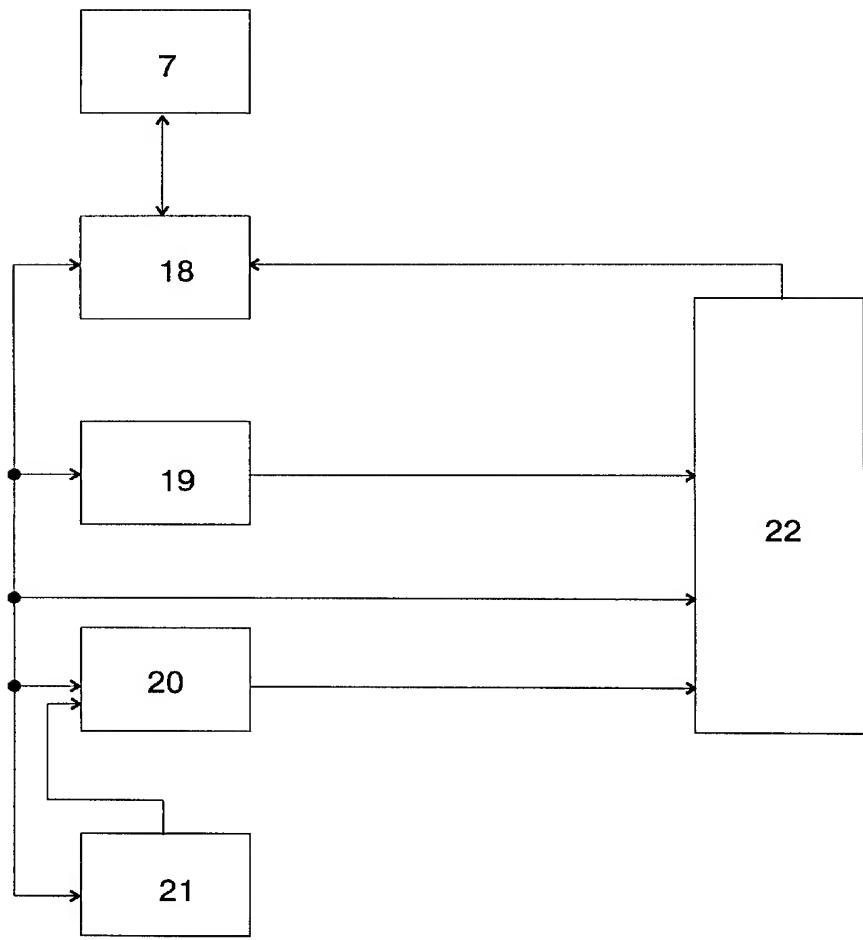
ФИГ. 2

R U ? 1 3 8 0 7 6 C 1



ФИГ. 3

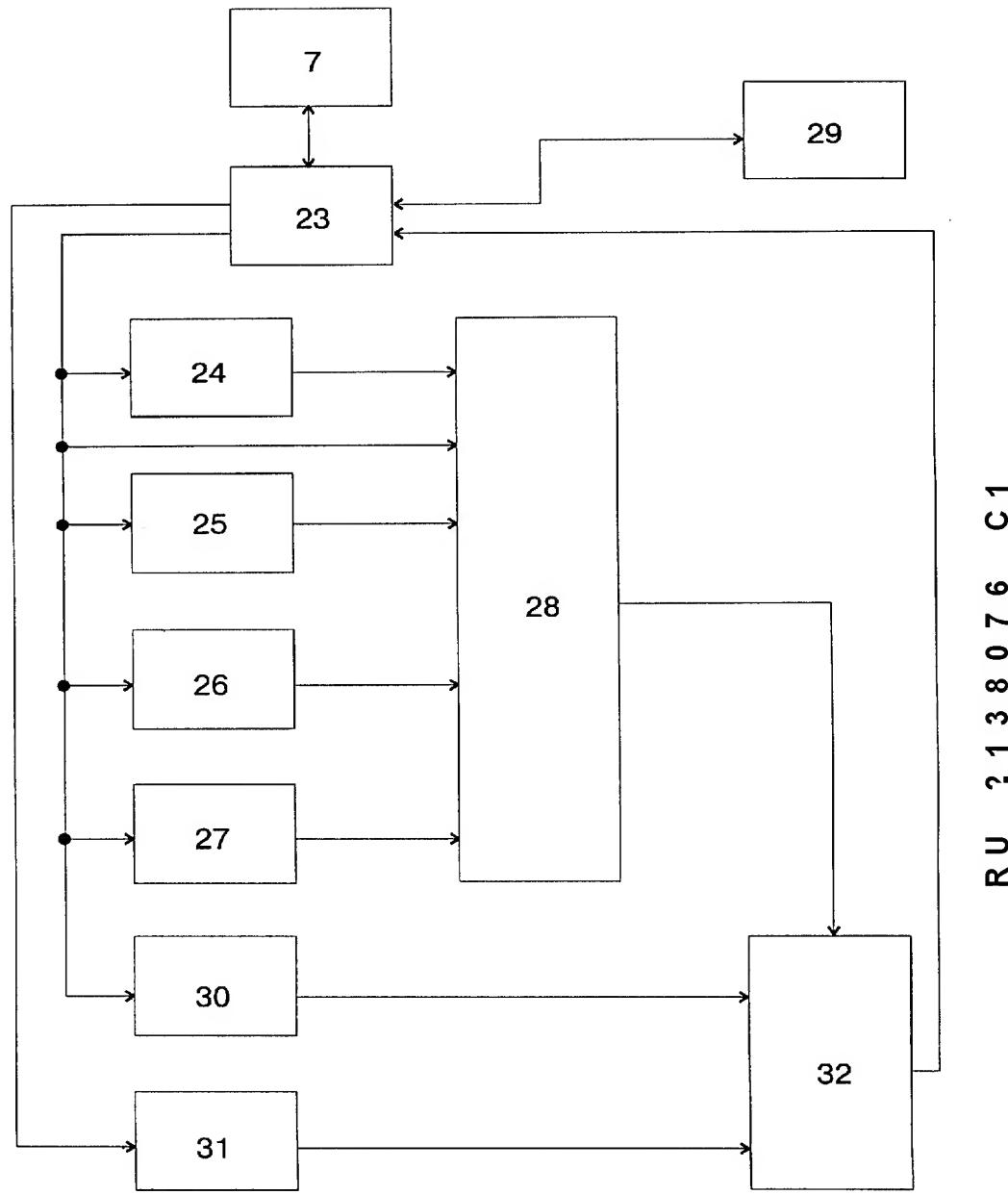
R U ? 1 3 8 0 7 6 C 1



Фиг.4

R U 2 1 3 8 0 7 6 C 1

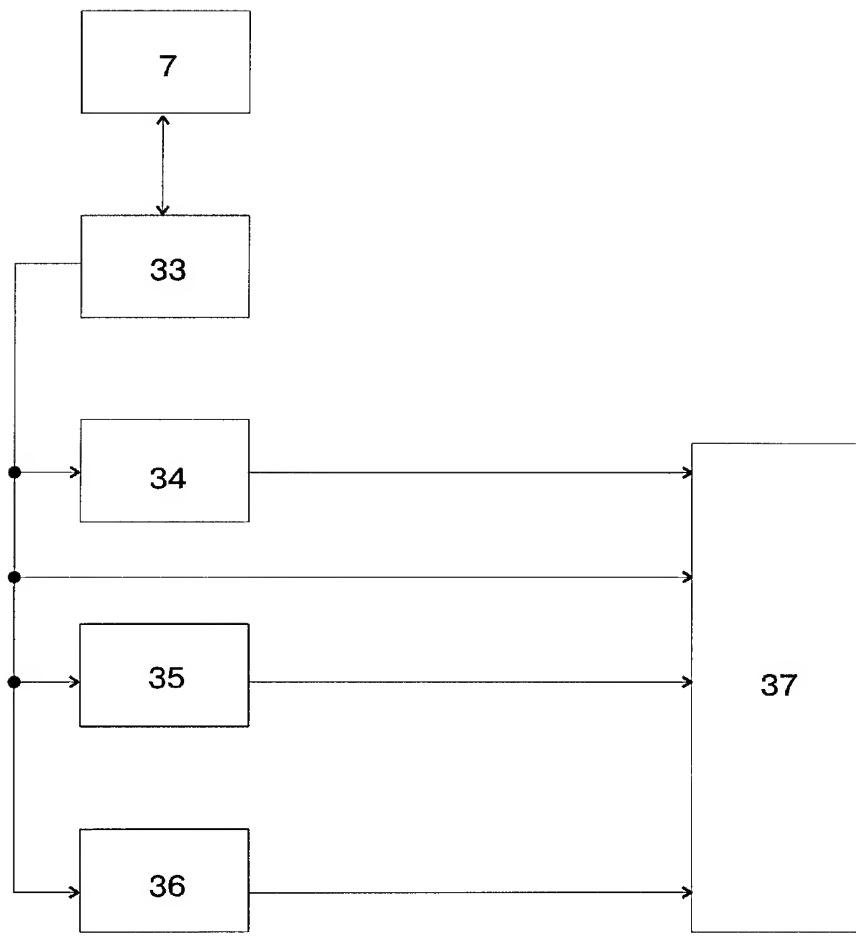
R U 2 1 3 8 0 7 6 C 1



ФИГ. 5

R U ? 1 3 8 0 7 6 C 1

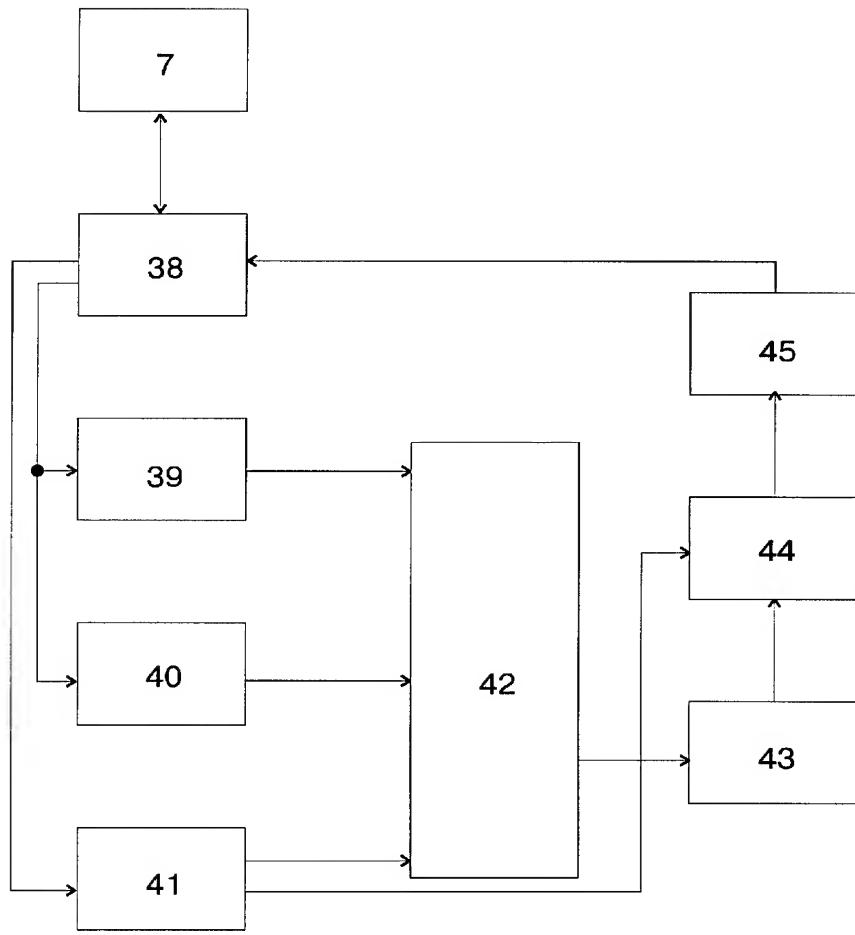
R U ? 1 3 8 0 7 6 C 1



ФИГ. 6

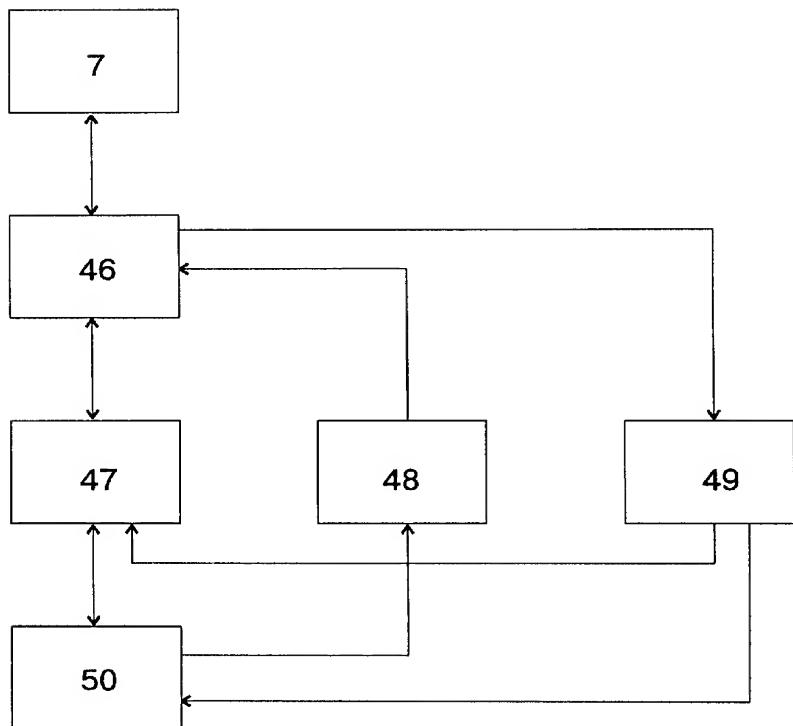
R U 2 1 3 8 0 7 6 C 1

R U ? 1 3 8 0 7 6 C 1



ФИГ. 7

R U 2 1 3 8 0 7 6 C 1



ФИГ. 8

R U ? 1 3 8 0 7 6 C 1

R U 2 1 3 8 0 7 6 C 1